

Peer-Review

Kundra, Sachin. 2026. "Crashmap360: A Data-Driven Infrastructure Recommendation System for Washington, D.C." *Journal of High School Science* 10 (3): 215–43. <https://doi.org/10.64336/001c.165268>.

Good work. However, you can extract and analyze much more data with much more use, effect and novelty than you have done in this manuscript. Please address my comments below.

[a] Tone down causal claims. You have no access to counterfactual analysis (e.g., before/after installation). There is no control for confounders (traffic volume, land use, road classification). There is no evidence that proximity to infrastructure caused lower severity.

[b] Add sensitivity and robustness analyses. Add a sensitivity analysis section showing how heuristic rankings change with different thresholds. Justify weights using literature or expert guidance, or normalize by injury counts.

[c] Improve severity scoring justification. The severity score combines frequency and severity, but this raises methodological concerns: Severity sums conflate "many mild crashes" with "few severe crashes." No normalization by exposure (traffic volume, pedestrian counts). A cluster with two fatalities can outrank one with many serious injuries. To correct these inconsistencies, Report frequency and severity separately alongside the composite score, Consider alternative metrics (e.g., fatalities per crash, weighted averages). Discuss ethical implications of weighting schemes.

[d] Include validation or baseline comparisons. There is no evaluation of whether the method outperforms simpler baselines. Missing comparisons: High-crash-frequency intersections without distance filtering. Kernel density heatmaps. Existing DDOT priority corridors. To correct this, Add a baseline comparison section. Show overlap (or lack thereof) with known high-risk corridors. Discuss whether Crashmap 360 provides new insights beyond crash density.

[e] Strengthen discussion of limitations and uncertainty. Be more cautious in interpreting proximity statistics. Acknowledge alternative explanations (placement bias, reporting bias). Provide opposing views and limitations of engineering policy (see paragraph at end of review).

[f] Expand depth to test and report whether some parameter pairs are more strongly associated with crashes than others at specific locations. This is a very important point. You can extract much more information and analysis out of your dataset than you are doing now which will increase the depth and value of your work. As written, your manuscript separates crashes into categories (night, speeding, bicycle, pedestrian), Analyzes each independently and identifies high-risk locations within each category. What it DOES NOT do (yet) is to compare categories against each other, quantify relative risk or disproportionate burden, test whether (for example) night disproportionately affects one crash type at certain locations or speeding and pedestrians are more likely to score more on severity than speeding and bicyclists or nighttime and bicyclists at certain locations. What you can do is create contingency tables (use severity, proportions or regression - not raw counts), then use chi-square or logistic regression to allow you to say (for example) "Nighttime is associated with a significantly higher odds of car-involved crashes relative to pedestrian or cyclist crashes at these locations". You can then say "At this location, nighttime car-involved crashes and speeding-related incidents dominate the observed crash profile, while pedestrian- and cyclist-involved crashes are comparatively rare. Consequently, interventions targeting vehicle speed and nighttime driving conditions may yield greater near-term safety benefits than pedestrian or bicycle infrastructure improvements."

This will elevate the novelty and usefulness of your work since What you will be describing will be both deeper and more novel than the current version of the manuscript, and it meaningfully elevates the work from "location recommendation" to "comparative risk prioritization". You will no longer be just mapping crashes. You will also be comparing competing safety needs at the same location and justifying why one intervention should take precedence. This is more novel, more useful, more analytical, more policy-relevant and more defensible as decision science.

You can then frame your novelty as a distance differentiator from current publications "Unlike prior approaches that analyze crash categories independently, this study introduces a comparative,

location-based framework that prioritizes infrastructure interventions based on the dominant crash mode and temporal context at each site.”

You already have the data for this. You just need to design the study. Ask your statistics teacher or chatgpt if you don’t know how. I look forward to your revised manuscript.

Include this in the discussion section:

“Although infrastructure safety modifications (e.g., traffic calming and built environment improvements) are intended to reduce crash frequency and severity, evidence from low-income and low-resource contexts is limited or mixed. Systematic reviews covering interventions in low- and middle-income countries have found few rigorous evaluations of infrastructure measures and suggest that legislation and enforcement currently show stronger impacts on crash outcomes than engineering alone. Moreover, studies show that built environment safety features do not always yield statistically significant reductions in crash rates in lower-income neighborhoods, and global trends indicate persistently higher crash burdens in these areas despite safety interventions. ”

<https://doi.org/10.3389/phrs.2022.1604499>

<https://doi.org/10.1016/j.trip.2022.100696>

As indicated in my earlier review, please provide a separate file describing verbatim my comments and how and where in the manuscript each of those comments was addressed.

Responses to Reviewer Comments

[a] Tone down causal claims. You have no access to counterfactual analysis (e.g., before/after installation). There is no control for confounders (traffic volume, land use, road classification). There is no evidence that proximity to infrastructure caused lower severity.

- The revised manuscript tones down causal claims throughout the paper. The Methods section and Limitations section now emphasize that there is no evidence that adding infrastructure in the identified locations would necessarily cause fewer severe crashes. The paper frames Crashmap 360 as a system for identifying candidate locations for further review, rather than as proof that missing infrastructure caused the crashes.

[b] Add sensitivity and robustness analyses. Add a sensitivity analysis section showing how heuristic rankings change with different thresholds. Justify weights using literature or expert guidance, or normalize by injury counts.

- The revised manuscript adds a sensitivity and robustness section in which different thresholds are tested, and ranking changes are evaluated using Spearman rank correlation coefficients and overlap percentages. This analysis is introduced in Section 2.8, reported in Section 3.5, and discussed in Section 4.4.

- The paper also justifies the thresholds more clearly, specifically for streetlights and speeding devices. This is discussed in Section 2.8. In addition, the threshold choices were discussed with a Vision Zero executive, as referenced in Section 2.8.

[c] Improve severity scoring justification. The severity score combines frequency and severity, but this raises methodological concerns: Severity sums conflate “many mild crashes” with “few severe crashes.” No normalization by exposure (traffic volume, pedestrian counts). A cluster with two fatalities can outrank one with many serious injuries. To correct these inconsistencies, Report frequency and severity separately alongside the composite score, Consider alternative metrics (e.g., fatalities per crash, weighted averages). Discuss ethical implications of weighting schemes.

- The revised manuscript discusses the ethical implications of weighting schemes in the severity scoring section. This discussion appears in Section 2.1.

- The paper also compares raw crash counts, or frequency, with severity sums using Spearman rank correlation coefficients. This analysis is introduced in Section 2.1, reported in Sections 3.1–3.3, and discussed in Sections 4.1–4.3.

- Traffic exposure is normalized using Annual Average Daily Traffic volumes in the Normalization by Exposure section. This is introduced in Section 2.5, with results reported in Sections 3.1–3.3 and discussed in Sections 4.1–4.3.

- Alternative metrics are also used in addition to the severity sum: severity average per crash, severe crash percentage, fatal crash percentage, and capped severity sum. These metrics are compared against each other and are displayed alongside the severity score. Frequency is also displayed under N_CRASHES. These additions are introduced in Section 2.1, reported in Sections 3.1–3.3, and discussed in Sections 4.1–4.3.

[d]Include validation or baseline comparisons. There is no evaluation of whether the method outperforms simpler baselines. Missing comparisons: High-crash-frequency intersections without distance filtering. Kernel density heatmaps. Existing DDOT priority corridors. To correct this, Add a baseline comparison section. Show overlap (or lack thereof) with known high-risk corridors. Discuss whether Crashmap 360 provides new insights beyond crash density.

- The revised manuscript compares results ranked by severity score and exposure-normalized results to the DDOT High Injury Network through overlap percentages in the top 5, top 10, and top 20 most prioritized locations. This comparison is introduced in Section 2.7, reported in Sections 3.1–3.3, and discussed in Sections 4.1–4.3.

The paper generally finds more overlap when the DDOT network is compared to results ranked by severity score than when it is compared to the exposure-normalized results. These results are described in more depth in the paper.

- The revised manuscript also compares clusters built from complete linkage to clusters built from DBSCAN, an unsupervised machine learning algorithm, to test how rankings change based on clustering method. This comparison is introduced in Section 2.6, reported in Sections 3.1–3.3, and discussed in Sections 4.1–4.3.

[e]Strengthen discussion of limitations and uncertainty. Be more cautious in interpreting proximity statistics. Acknowledge alternative explanations (placement bias, reporting bias). Provide opposing views and limitations of engineering policy (see paragraph at end of review).

- The revised manuscript narrows the final category-specific analysis by excluding pedestrian crashes because the available infrastructure layers did not allow a clean separation between sidewalks and crosswalks. This change improves methodological consistency by keeping the final analysis focused on crash categories with clearer infrastructure-proximity measures.

- The revised manuscript improves the limitations section by discussing placement bias, reporting bias, associational rather than causal relationships, and the limitations of engineering interventions. This discussion appears in Section 4.6.

- The revised manuscript acknowledges both opposing views and supporting views related to the methodology of the project in the literature review. This discussion appears in Section 1.1.

- One example of an opposing view is that, in lower-income countries, there was stronger evidence for legislation and enforcement than for roadway interventions.

- The revised manuscript adds more justification for performing analysis across different crash categories, normalizing by exposure, and looking beyond raw crash counts.

- The revised manuscript also introduces an alternative method to severity scoring, Equivalent Property Damage Only. The paper discusses why severity scoring is used instead of this method. This discussion appears in Section 2.1.

[f] Expand depth to test and report whether some parameter pairs are more strongly associated with crashes than others at specific locations. This is a very important point. You can extract much more information and analysis out of your dataset than you are doing now which will increase the depth and value of your work. As written, your manuscript separates crashes into categories (night, speeding, bicycle, pedestrian), Analyzes each independently and identifies high-risk locations within each category. What it DOES NOT do (yet) is to compare categories against each other, quantify relative risk or disproportionate burden, test whether (for example) night disproportionately affects one crash type at certain locations or speeding and pedestrians are more likely to score more on severity than speeding and bicyclists or nighttime and bicyclists at certain locations. What you can do is create contingency tables (use severity, proportions or regression - not raw counts), then use chi-square or logistic regression to allow you to say (for example) "Nighttime is associated with a significantly higher odds of car-involved crashes relative to pedestrian or cyclist crashes at these locations". You can then say "At this location, nighttime car-involved crashes and speeding-related incidents dominate the observed crash profile, while pedestrian- and cyclist-involved crashes are comparatively rare. Consequently, interventions targeting vehicle speed and nighttime driving conditions may yield greater near-term safety benefits than pedestrian or bicycle infrastructure improvements."

This will elevate the novelty and usefulness of your work since What you will be describing will be both deeper and more novel than the current version of the manuscript, and it meaningfully elevates the work from "location recommendation" to "comparative risk prioritization". You will no longer be just mapping crashes. You will also be comparing competing safety needs at the same location and justifying why one intervention should take precedence. This is more novel, more useful, more analytical, more policy-relevant and more defensible as decision science. You can then frame your novelty as a distinct differentiator from current publications "Unlike prior approaches that analyze crash categories independently, this study introduces a comparative, location-based framework that prioritizes infrastructure interventions based on the dominant crash mode and temporal context at each site."

You already have the data for this. You just need to design the study. Ask your statistics teacher or chatgpt if you don't know how. I look forward to your revised manuscript.

- Because crashes are filtered based on crash type and distance to relevant infrastructure before clustering, comparing those final candidate clusters to other crash types would produce bias. Instead of focusing on those filtered crash sets, the revised manuscript creates a broader crash-cluster dataset that does not filter based on crash type or distance to infrastructure.
- In this new dataset, four logistic regressions are used to examine which conditions and predictors are associated with different crash outcomes, including severity score ranking, high average severity ranking, severe injury percentage ranking, and exposure-normalized ranking.
- The predictors examined include the three different crash types, distance to infrastructure (streetlights, speeding devices, and bike lanes) and all three parameter pairs.
- Because this analysis uses a dataset different from the final candidate locations identified in the project, it focuses more on broader associations rather than location-specific comparisons. This broader analysis is more relevant to understanding general patterns across candidate locations without introducing bias from the original filtering process. This analysis is introduced in Section 2.9, reported in Section 3.6, and discussed in Section 4.5.

Thank you for addressing my comments. One methodological concern remains: the absence of a true validation framework demonstrating that Crashmap 360 provides better prioritization than simpler alternatives. While the manuscript now compares rankings generated

by severity scoring, exposure normalization, DBSCAN clustering, and the DDOT High Injury Network, these analyses primarily show that different methods produce different rankings rather than establishing that one approach is superior. To address this limitation, I recommend a temporal holdout validation. For example, use crash data from 2020–2022 to generate ranked intervention candidates and then evaluate those rankings against independent crash outcomes from 2023–2025. Crashmap 360 should be compared against simpler baselines such as raw crash frequency, severity score alone, exposure-normalized severity, and the DDOT High Injury Network. Performance metrics could include Precision@10, Recall@10, the proportion of future severe crashes captured, and the proportion of future fatalities captured within the highest-priority locations. Such an analysis would transform the manuscript from demonstrating a prioritization framework that produces different recommendations to demonstrating one that more accurately identifies future high-risk locations. Without an independent validation exercise of this nature, the manuscript can establish that Crashmap 360 generates alternative rankings, but it cannot yet demonstrate that those rankings are more effective or more informative than simpler prioritization approaches.

Reviewer Comment: the absence of a true validation framework demonstrating that Crashmap 360 provides better prioritization than simpler alternatives. While the manuscript now compares rankings generated by severity scoring, exposure normalization, DBSCAN clustering, and the DDOT High Injury Network, these analyses primarily show that different methods produce different rankings rather than establishing that one approach is superior. To address this limitation, I recommend a temporal holdout validation. For example, use crash data from 2020–2022 to generate ranked intervention candidates and then evaluate those rankings against independent crash outcomes from 2023–2025. Crashmap 360 should be compared against simpler baselines such as raw crash frequency, severity score alone, exposure-normalized severity, and the DDOT High Injury Network. Performance metrics could include Precision@10, Recall@10, the proportion of future severe crashes captured, and the proportion of future fatalities captured within the highest-priority locations. Such an analysis would transform the manuscript from demonstrating a prioritization framework that produces different recommendations to demonstrating one that more accurately identifies future high-risk locations. Without an independent validation exercise of this nature, the manuscript can establish that Crashmap 360 generates alternative rankings, but it cannot yet demonstrate that those rankings are more effective or more informative than simpler prioritization approaches.

[g] While a validation framework could be useful in determining which method is superior at predicting where car crashes will occur in the future, the framework does not fully align with the goals of Crashmap 360. Crashmap 360 does not claim to better predict where future car crashes may occur. Rather, it identifies locations where infrastructure gaps exist and where there is a pattern of car crashes related to a lack of specific infrastructure. While one can make the argument that the model is predicting where future car crashes occur, the main focus is where infrastructure could be added to address an abundance of car crashes. Thus, a comparison that determines a superior predictive method overlooks a key component of the project, the infrastructure focus. Because of these different objectives, one model should not be treated as superior to another. While one model may predict where a car crash will occur in the future more effectively than Crashmap 360, that model is not necessarily superior for identifying infrastructure-actionable locations.

[h] The project focus is further emphasized in the abstract and the literature review.

Fair enough. I concur with your rebuttal and your corresponding changes do bring the paper more in line with what you envisage.

However, I would like to request that you make the following changes:

1. explicitly state that the framework has not been independently validated against an external measure of intervention effectiveness and that future work should evaluate whether this proposed prioritization leads to better engineering decisions or safety outcomes than simpler prioritization methods.

2. Reinforce throughout the Discussion that infrastructure proximity is a screening criterion, not evidence of causation, and that recommended sites require subsequent engineering assessment before intervention.

3. Add a brief paragraph describing how agencies such as DDOT could operationally use the ranked candidate locations within an existing road safety review process.

4. The regression section feels out of place. Please make an effort to integrate this section into the manuscript. For example, verbiage such as “The ranking framework identifies locations requiring engineering review but does not reveal which crash characteristics are most strongly associated with highly severe clusters. Logistic regression was therefore used as an interpretability analysis to examine associations between cluster characteristics and high-severity outcomes.” and/or “The logistic regression analysis is not intended as an independent statistical exercise but as the second stage of the Crashmap 360 framework. After candidate infrastructure-review locations have been identified through clustering and ranking, regression analysis is used to determine which hotspot characteristics are independently associated with the most severe crash outcomes. These findings provide an additional layer of prioritization by helping distinguish candidate hotspots that exhibit combinations of characteristics associated with particularly severe crash patterns. Thus, Crashmap 360 combines spatial hotspot identification with statistical characterization of hotspot severity to support more informed infrastructure prioritization.”

Then, explicitly describe how its (the regression section’s) outputs modify, refine, or prioritize the hotspot recommendations.

5. Provide any code or associated math in a GitHub repository with a link to it in the manuscript. This is important for reproducibility of your work.

[i] Responses to Reviewer Comments

[j]

[k] explicitly state that the framework has not been independently validated against an external measure of intervention effectiveness and that future work should evaluate whether this proposed prioritization leads to better engineering decisions or safety outcomes than simpler prioritization methods.

- I now state that the framework has not had its effectiveness measured against other methods and that future work should explore this validation. (4.7)

[l] Reinforce throughout the Discussion that infrastructure proximity is a screening criterion, not evidence of causation, and that recommended sites require subsequent engineering assessment before intervention.

- I added this language along with other project limitations into the discussion section. I also emphasized the importance of additional review. (4.1, 4.2, 4.3, and 4.5)

[m] Add a brief paragraph describing how agencies such as DDOT could operationally use the ranked candidate locations within an existing road safety review process.

- I now discuss how agencies such as the DDOT could use Crashmap 360 alongside their High Injury Networks within an existing road safety review process. (4.6)

[n] The regression section feels out of place. Please make an effort to integrate this section into the manuscript. For example, verbiage such as “The ranking framework identifies locations

requiring engineering review but does not reveal which crash characteristics are most strongly associated with highly severe clusters. Logistic regression was therefore used as an interpretability analysis to examine associations between cluster characteristics and

[o] high-severity outcomes.” and/or “The logistic regression analysis is not intended as an independent statistical exercise but as the second stage of the Crashmap 360 framework. After candidate infrastructure-review locations have been identified through clustering and ranking, regression analysis is used to determine which hotspot characteristics are independently associated with the most severe crash outcomes. These findings provide an additional layer of prioritization by helping distinguish candidate hotspots that exhibit combinations of characteristics associated with particularly severe crash patterns. Thus, Crashmap 360 combines spatial hotspot identification with statistical characterization of hotspot severity to support more informed infrastructure prioritization.”

[p] Then, explicitly describe how its (the regression section’s) outputs modify, refine, or prioritize the hotspot recommendations.

[q] I now explain how the regression section fits into Crashmap 360 more clearly and emphasize how specific outputs prioritize the hotspot recommendations by identifying characteristics associated with high-severity outcomes. (2.9 and 4.5)

[r] Provide any code or associated math in a GitHub repository with a link to it in the manuscript. This is important for reproducibility of your work.

- A GitHub repository link is included at the bottom of the manuscript with all code used in the project. (Between 5. and 6.)

-

-

Thank you for addressing my comments. Accepted.